Contemporary Mathematics

Averaging Techniques for Reliable and Efficient *A Posteriori* Finite Element Error Control: Analysis and Applications

Carsten Carstensen and Max Jensen

Abstract. Local averaging techniques, which are used to postprocess discrete flux or stress approximations of low-order finite element schemes for elliptic boundary value problems, are applied for error control and adaptive mesh refinement. We put particular emphasis on the explicit calculation of all constants, arising in the proofs of reliability and efficiency, in terms of the known data and quantify the equivalence of local averaging techniques. We highlight and discuss a wide selection of applications for which averaging-based estimators provide highly accurate error control.

1. Introduction

In this section we present central concepts of *a posteriori* finite element error control. In Subsection 1.1 we give a brief introduction to reliability and efficiency of *a posteriori* error estimators. In Subsections 1.2 and 1.3 we consider error estimators as termination criteria for error control and in adaptive mesh-refining.

1.1. Reliability and Efficiency. To introduce error estimators we do not need to specify an underlying boundary value problem of the computation; in later sections, however, we will be concerned with differential equations

$$Lu = f,$$

paired with appropriate boundary conditions. Here L is in general an elliptic differential operator of second order, acting on functions with the domain Ω.

In the abstract framework it suffices to consider a function $p := p(\nabla u)$, which depends on the derivative of the unknown exact solution u, as well as the discrete counterpart $p_h := p(\nabla u_h)$, which depends on the known finite element solution u_h. Typically, p has the physical interpretation of a flux or stress.

Given a norm $\|\cdot\|$, the aim is to approximate the unknown error $\|p - p_h\|$ by a computable quantity η, called error estimator.

Definition 1.1. *(Error Estimator). A quantity η, which is thought of being an approximation to $\|p - p_h\|$, is called* a posteriori *error estimator, or estimator for brevity, if it is a computable function of known quantities such as $f, \Omega, \partial\Omega$ and u_h, p_h.*

The definition of error estimators is vague and not very useful on its own. The purpose of an estimator is to provide lower and upper error bounds. This is made precise by the terms *reliability* and *efficiency*. To introduce these two concepts, we consider a family of finite element meshes $(\mathcal{T}_h)_{h\in\mathcal{H}}$, where $h \in \mathcal{H}$ is a parameter representative for the mesh-size in $\mathcal{T}_h$. For quasi-uniform triangulations we assume that the index set $\mathcal{H}$ is a subset of $(0, \infty)$ and h is the diameter of the largest element.

2000 *Mathematics Subject Classification.* 65N15, 65N30.

DEFINITION 1.2. *(Reliability). An estimator η is called* reliable *if there is a constant $C_{\rm rel}$ and a bound*

$$\|p - p_h\| \le C_{\rm rel}\,\eta + \text{h.o.t.}_{\rm rel} \tag{1.1}$$

such that the function $\text{h.o.t.}_{\rm rel}$ *is* $o(\|p - p_h\|)$.

DEFINITION 1.3. *(Efficiency). An estimator η is called* efficient *if there is a constant $C_{\rm eff}$ and a bound*

$$\eta \le C_{\rm eff}\,\|p - p_h\| + \text{h.o.t.}_{\rm eff} \tag{1.2}$$

such that the function $\text{h.o.t.}_{\rm eff}$ *is* $o(\|p - p_h\|)$.

We say a function h.o.t. is $o(\|p - p_h\|)$ if, and only if, $\lim_{\mathcal{H}\ni h\to 0}\text{h.o.t.}/\|p - p_h\| = 0$.

DEFINITION 1.4. *(Asymptotic Exactness). An estimator is called* asymptotically exact *if it is reliable and efficient.*

We emphasize that $C_{\rm rel}$ and $C_{\rm eff}$ are multiplicative constants which do not depend on the mesh-size of the underlying finite element mesh $\mathcal{T}$. In the above definitions the abbreviation h.o.t. stands for 'higher-order terms'. In practical computations these terms are typically much smaller than η and $\|p - p_h\|$, but their size usually depends on the (unknown) smoothness of the exact solution or the (known) smoothness of the given data.

It is the mathematical task of *a posteriori* error analysis to provide sufficient and necessary conditions for the reliability and efficiency of error estimators as well as to characterise and estimate the constants $C_{\rm rel}$, $C_{\rm eff}$ and the higher-order terms $\text{h.o.t.}_{\rm rel}$, $\text{h.o.t.}_{\rm eff}$.

1.2. A Posteriori Error Control. There are at least two important areas where estimators are applied in practice: error estimation and adaptive mesh-refinement. For the first area, one is interested in a termination criterion for an algorithm of successively adapted mesh-refinements which guarantees that a given tolerance $\text{Tol} > 0$ is not exceeded:

$$\|p - p_h\| \le \text{Tol}.$$

Since the term $\|p - p_h\|$ is unknown, it is replaced by its upper bound (1.1) which leads to the criterion

$$C_{\rm rel}\eta + \text{h.o.t.}_{\rm rel} \le \text{Tol}. \tag{1.3}$$

For (1.3) to hold, it is evident that we require not only quantitative knowledge of η but also of $C_{\rm rel}$ and of $\text{h.o.t.}_{\rm rel}$. Notice that one may call the computable upper bound

$$\tilde\eta := C_{\rm rel}\eta + \text{h.o.t.}_{\rm rel}$$

an error estimator. Clearly, $\tilde\eta$ satisfies (1.1) with the reliability constant 1 and vanishing higher-order terms. If $\tilde\eta$ is not computable, the error control is incomplete and therefore useless. Fortunately, in the examples below, $\tilde\eta$ is computable and thus enables guaranteed error control.

Observe that the above error bound gives control only with respect to the norm $\|\cdot\|$ and not on other target functionals. In this paper we focus on energy norms and so ignore goal-oriented error control. The latter is also important and the reader is referred to [**1, 3**], where arguments are detailed on how to reduce the more general problem to energy-norm control, and to [**7**] for a survey of the work of Rannacher et al. on a very successful computational approach.

1.3. Adaptive Mesh-Refining. Error estimators usually not only give a bound on the global error, but also contain information on the local approximation quality of p_h to p. For instance, η can take the form

$$\eta^2 = \sum_{T\in\mathcal{T}} \eta_T^2, \tag{1.4}$$

where η_T are computable elementwise contributions to η. The second area of application of error estimators relies on the observation. One then interprets η_T as a local *indicator* and refines the

element T if the associated value η_T is relatively large. For example, one can flag all elements T for refinement which satisfy the inequality

$$1/2 \max\{\eta_K : K \in \mathcal{T}\} \leq \eta_T. \tag{1.5}$$

Notice that further strategies may be required to avoid hanging nodes and degenerated elements. The use of η_T as a refinement criterion is often based on heuristics; one should then speak of a *refinement indicator* η_T and not of an error indicator.

Notice that $C_{\mathrm{rel}}, C_{\mathrm{eff}}$ do not enter in (1.5) while $\mathrm{h.o.t.}_{\mathrm{rel}}, \mathrm{h.o.t.}_{\mathrm{eff}}$ are simply ignored.

The rigorous justification of adaptive mesh-refining algorithms started with [**19, 20, 22**] by a proof of error-reduction properties. In these publications, even the *rate* of convergence of the numerical approximations to the exact solution is specified. However, the number of elements generated by the considered adaptive algorithms is not controlled, which may affect the performance of the method. In [**8**] coarsening steps are introduced to prove optimality properties in respect to the performance of the adaptive mesh-refinement. However, coarsening seems to be, for elliptic problems, theoretically motivated in first place and appears to be unnecessary in practice.

2. Preliminaries and Notation

Let Ω be a bounded Lipschitz domain in $\mathbb{R}^2$ with a piecewise affine boundary Γ. Assume that Ω is exactly covered by a mesh $\mathcal{T}$, i.e. $\cup\mathcal{T} = \overline{\Omega}$. We only consider closed triangular elements $T \in \mathcal{T}$.

The vertices $a, b, c \in \mathbb{R}^2$ of the element $T = \mathrm{conv}\{a, b, c\}$ are called nodes; $\mathcal{N}$ denotes the set of all nodes. We let $\mathcal{N}_\Gamma := \Gamma \cap \mathcal{N}$ be the set of exterior and $\mathcal{N}_\Omega := \mathcal{N} \setminus \mathcal{N}_\Gamma$ be the set of interior nodes. We write $\mathcal{E}$ for the set of edges E; $\mathcal{E}_\Omega$ denotes the interior edges and $\mathcal{E}_\Gamma := \{E \in \mathcal{E} : E \subset \Gamma\} = \mathcal{E}\backslash\mathcal{E}_\Omega$ denotes the boundary edges. Intersecting distinct elements share either one vertex or an edge. Hanging nodes are excluded for the ease of the presentation. For each node $z \in \mathcal{N}$ let $\mathcal{E}_z := \{E \in \mathcal{E} : z \in E \cap \mathcal{N}\}$ and $\mathcal{T}_z := \{T \in \mathcal{T} : z \in T \cap \mathcal{N}\}$. To each edge E a unit normal vector ν_E with fixed orientation is associated; if $E \subseteq \partial\Omega$, set $\nu_E = \nu$, the outer unit normal along $\partial\Omega$. The length of $E \in \mathcal{E}$ is denoted by $h_E = \mathrm{diam}(E)$ and the diameter of $\mathcal{T}_z$, $z \in \mathcal{N}$, by $h_{\mathcal{T}_z}$. Similarly, the diameter of an element T is denoted by h_T. The area of $T \in \mathcal{T}$ is abbreviated by $|T| = \mathcal{L}^2(T)$. Here $\mathcal{L}^2(T)$ is the two-dimensional Lebesgue measure of T.

We let $P_k(T)$ be the set of algebraic polynomials with total degree less than or equal to k. Furthermore, $P_k(T, \mathbb{M})$ is the set of $\mathbb{M}$-valued elements, where $\mathbb{M}$ is a space of real scalars, vectors or matrices. We set

$$P_k(\mathcal{T}, \mathbb{M}) := \{v_h \in L^\infty(\Omega) : \forall\, T \in \mathcal{T},\, v_h|_T \in P_k(T, \mathbb{M})\},$$
$$P_k(\mathcal{T}) := P_k(\mathcal{T}, \mathbb{R}).$$

We let $\mathcal{P}_h$ be a space which contains all possible values of the p_h we introduced in the previous section. We make hereby the assumption that $\mathcal{P}_h$ is a subspace of $P_1(\mathcal{T}, \mathbb{M})$. Typically, $\mathcal{P}_h = P_0(\mathcal{T}, \mathbb{M})$ or $\mathcal{P}_h = P_1(\mathcal{T}, \mathbb{M})$. However, in principle, also other choices of $\mathcal{P}_h$ are admissible; for instance, to take the effect of the boundary conditions into account.

We shall see in the subsequent sections that the size of the error $\|p - p_h\|$ can be estimated by comparing p_h with functions in the space

$$\mathcal{Q}_h := \{v_h \in P_1(\mathcal{T}, \mathbb{M}) : v_h \text{ continuous}\}.$$

Given a subset ω of Ω, we abbreviate

$$\mathcal{P}_h|_\omega := \{p_h|_\omega : p_h \in \mathcal{P}_h\} \quad \text{and} \quad \mathcal{Q}_h|_\omega := \{q_h|_\omega : q_h \in \mathcal{Q}_h\}.$$

The nodal basis functions $\varphi_z \in P_1(\mathcal{T})$, $z \in \mathcal{N}$, are defined by

$$\varphi_z(z') = \begin{cases} 1 : z = z' \\ 0 : z \neq z' \end{cases}, \qquad z' \in \mathcal{N}.$$

Consequently,

$$0 \leq \varphi_z \leq 1, \quad \mathrm{supp}\,\varphi_z = \mathcal{T}_z, \quad \text{and} \quad \textstyle\sum_{z\in\mathcal{N}} \varphi_z = 1.$$

2.1. Averaging Operators. We construct averaging operators $A : \mathcal{P}_h \to \mathcal{Q}_h$ from local linear operators $A_z : P_1(\mathcal{T}_z;\mathbb{M}) \to \mathbb{M}$ with the formula

$$A(p_h) := \sum_{z\in\mathcal{N}} A_z(p_h|_{\mathcal{T}_z})\,\varphi_z, \qquad p_h \in \mathcal{P}_h. \tag{2.1}$$

We assume that the local averaging operators A_z are *preserving*, that is

$$A_z(f) = f(z) \tag{2.2}$$

for all $f \in N(\mathcal{T}_z) := (\mathcal{P}_h|_{\mathcal{T}_z}) \cap (\mathcal{Q}_h|_{\mathcal{T}_z})$.

Example 2.1 (Averaging of Nodal Values). *Amongst the easiest local averaging techniques is the averaging of the function values of p_h at a node $z \in \mathcal{N}$; that is, A_z is defined as weighted mean of all the different values $(v|_T)(z)$ where $z \in T$:*

$$A_z(v) := \sum_{T\in\mathcal{T}_z} \lambda_{z,T}\, v|_T(z) \quad \textit{for all } v \in P_1(\mathcal{T}_z;\mathbb{M}),\ v \in \mathcal{N}. \tag{2.3}$$

Here the $\lambda_{z,T}$ are real coefficients for which condition (2.2) *is equivalent to*

$$\forall\, z \in \mathcal{N} : \textstyle\sum_{T\in\mathcal{T}_z} \lambda_{z,T} = 1. \tag{2.4}$$

Example 2.2 (ZZ Averaging). *The special case*

$$\lambda_{z,T} := |T|/|\mathcal{T}_z| \quad \textit{for all } T \in \mathcal{T},\ z \in \mathcal{N}, \tag{2.5}$$

where $|\cdot|$ denotes the area, is our interpretation of a gradient recovery. The case $\mathcal{P}_h = \mathcal{P}_0(\mathcal{T})$ is due to Zienkiewicz and Zhu [**23**]. *The corresponding operator $Z := A$ with $A_z := Z_z$ reads*

$$Z(p_h) = \sum_{z\in\mathcal{N}} \Big(\sum_{T\in\mathcal{T}_z} |T|/|\mathcal{T}_z|(p_h|_T)(z) \Big)\, \varphi_z.$$

Since the choice (2.5) *immediately implies* (2.4)*, condition* (2.2) *is satisfied.*

2.2. Estimators. We define, for any fixed $p_h \in \mathcal{P}_h$, the averaging estimators

$$\eta_M := \min_{r_h\in\mathcal{Q}_h} \|p_h - r_h\|_{L^2(\Omega,\mathbb{M})} \le \eta_A := \|p_h - A(p_h)\|_{L^2(\Omega,\mathbb{M})}.$$

For the ZZ averaging operator from Example 2.2, we introduce

$$\eta_Z := \|p_h - Z(p_h)\|_{L^2(\Omega,\mathbb{M})}.$$

Finally, given any $p_h \in \mathcal{P}_h$ and $E \in \mathcal{E}$, let $[p_h]|_E$ denote the jump of p_h across the edge E with the L^2-norm $\|[p_h]|_E\|_{L^2(E)}$ along E; then,

$$\eta_{\mathcal{E}} := \Big(\sum_{E\in\mathcal{E}} h_E \|[p_h]|_E\|^2_{L^2(E)} \Big)^{1/2}.$$

We introduce the convention $[p_h]|_E := 0$ for boundary edges $E \in \mathcal{N}_\Omega$ to unify the treatment of interior and exterior edges in some of the forthcoming proofs.

3. Efficiency

There is no need to specify a particular boundary value problem in order to analyze the efficiency of the above estimators. In this section we prove that averaging estimators are efficient for *any* problem which concerns the approximation of smooth functions by elements in $\mathcal{P}_h$.

We show first that η_M is efficient in the sense of

$$\eta_M \le \|p - p_h\| + \text{ h.o.t.} \tag{3.1}$$

The surprising fact is that p can be *any* smooth function, e.g. $p \in H^1(\Omega;\mathbb{R}^n)$, and $\|\cdot\|$ *any* norm such that

$$\min_{q_h\in\mathcal{Q}_h} \|p - q_h\| = o(\|p - p_h\|). \tag{3.2}$$

The proof of (3.1) only requires the triangle inequality and is therefore applicable to a wide range of applications: for any $q_h \in \mathcal{Q}_h$ we have

$$\eta_M \le \|p_h - q_h\| \le \|p - p_h\| + \|p - q_h\|.$$

Thus, efficiency (3.1) holds without any reference to the underlying boundary value problem. For instance, if p_h denotes some flux approximation in a first-order conforming or nonconforming or lowest-order mixed FEM, then the L^2-error $\|p-p_h\|$ is of first order, while (3.2) is of second order provided the exact flux p is sufficiently smooth.

We now show that η_M and η_A are equivalent up to multiplicative constants.

THEOREM 3.1. *Let A be an averaging operator defined as in* (2.1) *with preserving local averaging operators A_z. Moreover, we assume that the A_z are uniformly bounded in the following sense*

$$\exists\, C_{\mathrm{uni}} > 0\, \forall\, z \in \mathcal{N}\, \forall\, p_z \in P_1(\mathcal{T}_z, \mathbb{M})\ : C_{\mathrm{uni}}\, \|p_z\|_{L^2(\mathcal{T}_z,\mathbb{M})} \ge \|A_z p_z\|_{\mathbb{M}}\, \sqrt{|\mathcal{T}_z|}. \tag{3.3}$$

Then

$$\eta_M \le \eta_A \le (1 + \sqrt{3/2}\, C_{\mathrm{uni}})\, \eta_M.$$

for all $p_h \in \mathcal{P}_h$.

PROOF. The first inequality is obvious and the proof concerns the second. The operator norm of A is bounded by all $\sqrt{3/2}\, C_{\mathrm{uni}}$ because for all $p_h \in P_1(\mathcal{T}, \mathbb{M})$

$$\begin{aligned}
\|Ap_h\|^2_{L^2(\Omega,\mathbb{M})} &= \int_\Omega \Big\| \sum_{z\in\mathcal{N}} A_z(p_h|_{\mathcal{T}_z})\, \varphi_z \Big\|^2_{\mathbb{M}} dx \le \int_\Omega 3 \sum_{z\in\mathcal{N}} \|A_z(p_h|_{\mathcal{T}_z})\|^2_{\mathbb{M}}\, \varphi_z^2\, dx \\
&\le 3 \sum_{z\in\mathcal{N}} \frac{C^2_{\mathrm{uni}}\, \|p_h\|^2_{L^2(\mathcal{T}_z,\mathbb{M})}}{|\mathcal{T}_z|} \int_\Omega \varphi_z^2\, dx = 3 \sum_{z\in\mathcal{N}} \frac{C^2_{\mathrm{uni}}\, \|p_h\|^2_{L^2(\mathcal{T}_z,\mathbb{M})}}{|\mathcal{T}_z|}\, \frac{|\mathcal{T}_z|}{6} \\
&= 3/2\, C^2_{\mathrm{uni}} \|p_h\|^2_{L^2(\Omega,\mathbb{M})}.
\end{aligned}$$

where we used $\|\varphi_z\|^2_{L^2(\mathcal{T}_z)} = |\mathcal{T}_z|/6$ for $T \in \mathcal{T}_z$. Now let $p_h \in \mathcal{P}_h$. There is a unique decomposition of p_h into a component p_c in $\mathcal{Q}_h$ and into a component p_d in its orthogonal complement in $L^2(\Omega;\mathbb{M})$. Because the local averaging operators A_z are preserving, we have $Ap_c = p_c$. Thus

$$\|p_h - Ap_h\|_{L^2(\Omega,\mathbb{M})} = \|p_d - Ap_d\|_{L^2(\Omega,\mathbb{M})} \le \|p_d\|_{L^2(\Omega,\mathbb{M})} + \|Ap_d\|_{L^2(\Omega,\mathbb{M})} \le (1 + \sqrt{3/2}\, C_{\mathrm{uni}})\, \eta_M,$$

where we used that $\eta_M = \|p_d\|_{L^2(\Omega)}$. □

Recall the averaging of nodal values discussed in Example 2.1. Computing the minimum of the polynomial

$$(\alpha, \beta) \mapsto \int_0^1 \int_0^{1-y} (\alpha\, x + \beta\, y + 1)^2\, dx\, dy$$

and scaling show that $\|p_z|_T(z)\|_{\mathbb{M}}$ is bounded by $6\, \|p_z\|_{L^2(T,\mathbb{M})}/\sqrt{|T|}$. Hence for $p_z \in P_1(\mathcal{T}_z, \mathbb{M})$

$$\sqrt{|\mathcal{T}_z|}\, \|A_z p_z\|_{\mathbb{M}} \le 6 \sum_{T\in\mathcal{T}_z} |\lambda_{T,z}|\, \frac{\sqrt{|\mathcal{T}_z|}}{\sqrt{|T|}}\, \|p_z\|_{L^2(T,\mathbb{M})} \le 6 \Big(\sum_{T\in\mathcal{T}_z} |\lambda_{T,z}|^2\, \frac{|\mathcal{T}_z|}{|T|} \Big)^{1/2} \|p_z\|_{L^2(\mathcal{T}_z,\mathbb{M})}.$$

Consequently,

$$C_{\mathrm{uni}} = \max_{z\in\mathcal{N}} 6 \Big(\sum_{T\in\mathcal{T}_z} |\lambda_{T,z}|^2\, \frac{|\mathcal{T}_z|}{|T|} \Big)^{1/2}.$$

For the ZZ estimator we conclude

$$C_{\mathrm{uni}} = \max_{z\in\mathcal{N}} 6 \Big(\sum_{T\in\mathcal{T}_z} \frac{|\mathcal{T}_z|^2}{|T|^2}\, \frac{|\mathcal{T}_z|}{|T|} \Big)^{1/2} = 6.$$

Notice that we do not require shape regularity.

REMARK 3.1. *In* [**10**] *it is shown that the above bound can be sharpened with a more elaborate argument based on Ascoli's lemma and the explicit calculation of stiffness matrices. There it is demonstrated that*

$$\eta_Z \leq \sqrt{10}\,\eta_M \approx 3.2\,\eta_M \leq (1+\sqrt{3/2}\,6)\,\eta_M \approx 8.3\,\eta_M.$$

In this paper also the three dimensional setting is analyzed.

We now turn to the equivalence of η_M and $\eta_{\mathcal{E}}$.

THEOREM 3.2. *Suppose that $\mathcal{T}$ is shape regular. Then there is a constant $C_{\rm sr}$ such that*

$$h_E^2 \leq C_{\rm sr}\,|T|$$

for all $T \in \mathcal{T}$ and edges E of T. There also is an integer constant $C_{\rm en}$ which denotes the maximum number of elements which share the same node. The estimators η_M, η_Z and $\eta_{\mathcal{E}}$ then satisfy the relationship

$$(C_{\rm en}\,C_{\rm sr})^{-1/2}\,\eta_Z \leq \eta_{\mathcal{E}} \leq (24\,C_{\rm sr})^{1/2}\,\eta_M.$$

PROOF. Suppose that the jumps in the patch $\mathcal{T}_z$ are computed in clockwise direction, i.e. given an edge $E \in \mathcal{E}_z$ which is the boundary between the elements $T_1, T_2 \in \mathcal{T}_z$, we set $[p_h]|_E = p_h|_{T_1} - p_h|_{T_2}$ if T_1 is clockwise positioned ahead of T_2 with respect to z. Then

$$\sum_{E\in\mathcal{E}_z} [p_h]|_E(z) = 0.$$

Thus, for all $T \in \mathcal{T}_z$, we have

$$\|p_h|_T(z) - Z(p_h|_{\mathcal{T}_z})\|_{\mathbb{M}} \leq \max_{S\in\mathcal{T}_z} \|p_h|_T(z) - p_h|_S(z)\|_{\mathbb{M}} \leq \frac{1}{2}\sum_{E\in\mathcal{E}_z} \|[p_h]|_E(z)\|_{\mathbb{M}}.$$

For all affine functions p on $E \in \mathcal{E}_z$ the bound

$$h_E \|p(z)\|_{\mathbb{M}}^2 \leq 4\,\|p\|_{L^2(E,\mathbb{M})}^2$$

holds. Thus

$$\begin{aligned}
\eta_Z^2 &= \int_\Omega \Big\| \sum_{z\in\mathcal{N}} (p_h - Z(p_h|_{\mathcal{T}_z}))\,\varphi_z \Big\|_{\mathbb{M}}^2 dx \leq 3\int_\Omega \sum_{z\in\mathcal{N}} \|p_h - Z(p_h|_{\mathcal{T}_z})\|_{\mathbb{M}}^2\,\varphi_z^2\,dx \\
&\leq 3\int_\Omega \sum_{z\in\mathcal{N}} \Big(\frac{1}{2}\sum_{E\in\mathcal{E}_z} \|[p_h]|_E(z)\|_{\mathbb{M}}\Big)^2 \varphi_z^2\,dx \leq \frac{3}{4}\int_\Omega \sum_{z\in\mathcal{N}} C_{\rm en} \sum_{E\in\mathcal{E}_z} \|[p_h]|_E(z)\|_{\mathbb{M}}^2\,\varphi_z^2\,dx \\
&= \frac{3\,C_{\rm en}}{24} \sum_{z\in\mathcal{N}} \sum_{E\in\mathcal{E}_z} \|[p_h]|_E(z)\|_{\mathbb{M}}^2\,|\mathcal{T}_z| \leq \frac{C_{\rm en}\,C_{\rm sr}}{8} \sum_{z\in\mathcal{N}} \sum_{E\in\mathcal{E}_z} \|[p_h]|_E(z)\|_{\mathbb{M}}^2\,h_E^2 \leq C_{\rm en}\,C_{\rm sr}\,\eta_{\mathcal{E}}^2,
\end{aligned}$$

taking for the last inequality into account that the double sum passes over each edge twice. With p_d like in the proof of the last theorem, the chain of inequalities

$$\begin{aligned}
\eta_{\mathcal{E}}^2 &= \sum_{E\in\mathcal{E}} h_E \int_E \Big\| \sum_{z\in\mathcal{N}} [p_d]|_E(z)\,\varphi_z \Big\|_{\mathbb{M}}^2 dx \leq \sum_{z\in\mathcal{N}} \sum_{E\in\mathcal{E}_z} h_E\,\|[p_d]|_E(z)\|_{\mathbb{M}}^2\,\|\varphi_z\|_{L^2(E)}^2 \\
&= \sum_{z\in\mathcal{N}} \sum_{E\in\mathcal{E}_z} \frac{h_E^2}{3}\,\|[p_d]|_E(z)\|_{\mathbb{M}}^2 \leq \sum_{z\in\mathcal{N}} \sum_{T\in\mathcal{T}_z} \frac{2\,h_E^2}{3}\,\|p_d|_T(z)\|_{\mathbb{M}}^2 \leq \sum_{z\in\mathcal{N}} \sum_{T\in\mathcal{T}_z} \frac{2\,h_E^2}{3}\,\frac{36\,\|p_d\|_{L^2(T,\mathbb{M})}^2}{|T|} \\
&\leq 24\,C_{\rm sr}\,\eta_M^2
\end{aligned}$$

completes the argument. □

4. Local Approximation Operators

The averaging operators A, considered in the last section, belong to the important class of the local approximation operators. By local approximation operators we mean operators of the structure

$$J : V \to V_h,\ p \mapsto \sum_{z\in\mathcal{N}} ((\Pi_z \circ A_z)(p|_{\mathcal{T}_z}))\,\varphi_z. \tag{4.1}$$

The operators J have through V, V_h and Π_z additional flexibility compared to A, which is used to incorporate additional constraints to the approximation if needed. For example, one may choose V, V_h and Π_z such that $J(v)$ satisfies the boundary conditions of the underlying boundary value problem exactly. However, also in other contexts constraints arise which can be incorporated into J; for instance, obstacle conditions in the field of variational inequalities. To give consideration to a wide range of applications, we cover the general case.

To be compatible with the framework of the operators A, the space V needs to contain, in general, the set $\mathcal{P}_h$. In other situations local approximation operators are used to project the exact solution of the boundary value problem into the finite element space. For these applications V needs to contain functions of $H^1(\Omega)$-type. For this reason we assume that V is a subset of the broken Sobolev space

$$H^1(\mathcal{T},\mathbb{M}) := \{v \in L^2(\Omega,\mathbb{M}) : \forall T \in \mathcal{T} : v|_T \in H^1(T,\mathbb{M})\}.$$

The space $H^1(\mathcal{T},\mathbb{M})$ is equipped with the semi-norm

$$|p|^2_{H^1(\mathcal{T},\mathbb{M})} := \sum_{T\in\mathcal{T}} |(p|_T)|^2_{H^1(T,\mathbb{M})} + \sum_{E\in\mathcal{E}_\Omega} \|[p]|_E\|^2_{L^2(E,\mathbb{M})}.$$

and the norm

$$\|p\|^2_{H^1(\mathcal{T},\mathbb{M})} := \|p\|_{L^2(\Omega,\mathbb{M})} + |p|^2_{H^1(\mathcal{T},\mathbb{M})}.$$

We now turn to V_h. For all $z \in \mathcal{N}$ let $\mathbb{A}_z$ be a non-empty, convex and closed subset of $\mathbb{M}$ and let $\Pi_z : \mathbb{M} \to \mathbb{A}_z$ be the orthogonal projection onto $\mathbb{A}_z$ in the canonical scalar product of $\mathbb{M}$. We assume the $\mathbb{A}_z$ are chosen such that the following compatibility condition between V and V_h is satisfied:

$$\exists\, C_{\rm pf} > 0\, \forall\, z \in \mathcal{N}\, \forall\, p \in V|_{\mathcal{T}_z} : \|p - \Pi_z\bar{p}\|_{L^2(\mathcal{T}_z,\mathbb{M})} \le C_{\rm pf}\, |h\, p|_{H^1(\mathcal{T}_z,\mathbb{M})}, \tag{4.2}$$

where

$$h : \Omega \to \mathbb{R}, x \in T \mapsto \operatorname{diam}(T) \text{ and } \bar{p} := \fint_{\mathcal{T}_z} p\, dx.$$

Here and throughout the text we identify real numbers, like $\Pi_z\bar{p}$ in the above formula, with the constant function attaining this function value. In our analysis, (4.2) is used similar to a Poincaré-Friedrichs inequality. If (4.2) is fulfilled we define

$$V_h := \Big\{ \sum_{z\in\mathcal{N}} a_z\, \varphi_z : a_z \in \mathbb{A}_z \Big\} \subset P_1(\mathcal{T},\mathbb{M}).$$

Instead of the orthogonal projection other choices for Π_z are possible. We remark that in some of the subsequent proofs the expansion factor of the projection has to be included explicitly for more general Π_z. For the orthogonal projection the factor is equal to 1:

$$\forall p_1, p_2 \in \mathbb{M} : \|\Pi_z p_1 - \Pi_z p_2\|_{\mathbb{M}} \le \|p_1 - p_2\|_{\mathbb{M}}. \tag{4.3}$$

We now turn to the operators $A_z : V_h|_{\mathcal{T}_z} \to \mathbb{M}$. We assume they satisfy the condition that

$$\exists\, C_{\rm j} > 0\, \forall\, z \in \mathcal{N}\, \forall\, p \in V|_{\mathcal{T}_z} : \|A_z p - \bar{p}\|_{L^2(\mathcal{T}_z,\mathbb{M})} \le C_{\rm j}\, |h\, p|_{H^1(\mathcal{T}_z,\mathbb{M})}. \tag{4.4}$$

We show that the ZZ estimator satisfies (4.4).

THEOREM 4.1. *Suppose that $\mathcal{T}$ is shape-regular with $C_{\rm sr}$ and $C_{\rm en}$ like in Theorem 3.2. Then*

$$\|Zp - \bar{p}\|_{L^2(\mathcal{T}_z,\mathbb{M})} \le \max\{\sqrt{C_{\rm en}\, C_{\rm sr}}, C_{\rm p}\}\, |h\, p|_{H^1(\mathcal{T}_z,\mathbb{M})}$$

where $C_{\rm p}$ is the constant in the Poincaré inequality for the elements of the triangulation $\mathcal{T}$.

PROOF. We have

$$\begin{aligned}\|Zp-\bar p\|_{L^2(\mathcal{T}_z,\mathbb{M})} &\le \|p-Zp\|_{L^2(\mathcal{T}_z,\mathbb{M})}+\|p-\bar p\|_{L^2(\mathcal{T}_z,\mathbb{M})}\\ &\le \sqrt{C_{\rm en}\,C_{\rm sr}}\,\eta_{\mathcal{E}}+\textstyle\sum_{T\in\mathcal{T}_z} C_{\rm p}\,\mathrm{diam}(T)\,|p|_{H^1(T,\mathbb{M})}\\ &\le \max\{\sqrt{C_{\rm en}\,C_{\rm sr}},C_{\rm p}\}\,|h\,p|_{H^1(\mathcal{T}_z,\mathbb{M})},\end{aligned}$$

according to Theorem 3.2. □

The A_z are not necessarily pointwise interpolation operators. It is self-evident that

$$A_z\,p=\bar p \tag{4.5}$$

satisfies (4.4). We remark that similarly the important class of quasi-interpolation operators introduced by Clément fits into the framework of local approximation operators, provided the definition of A_z is relaxed to $A_z:\Omega_z\to\mathbb{R}$, where Ω_z is a superset of $\mathcal{T}_z$. While in the interior of the domain Clément operators also utilise (4.5), for boundary nodes z the average of p in $\mathcal{T}_z$ is not assigned to the coefficient of φ_z but added to the coefficient of a neighbouring interior node.

In the next theorem we consider the rate of convergence of local approximation operators.

THEOREM 4.2. *Local approximation operators J, satisfying* (4.2) *and* (4.4)*, are locally first-order approximating, which means there is a constant $C>0$ such that for all $p\in V$*

$$\|p-Jp\|_{L^2(\Omega,\mathbb{M})}\le C|h\,p|_{H^1(\mathcal{T},\mathbb{M})}.$$

For J the constant C is equal to $\sqrt{3}\,(C_{\rm pf}+C_{\rm j})$. If $\mathcal{T}$ is shape-regular, then all $p\in V$ also satisfy

$$\|h^{-1}\,(p-Jp)\|_{L^2(\Omega,\mathbb{M})}\le\sqrt{3}\,C_{\rm t}\,(C_{\rm pf}+C_{\rm j})\,|p|_{H^1(\mathcal{T},\mathbb{M})},$$

where

$$C_{\rm t}=\max_{z\in\mathcal{N}}\max_{S,T\in\mathcal{T}_z}\frac{\mathrm{diam}(S)}{\mathrm{diam}(T)}.$$

PROOF. Given $p\in V$ let $p_z:=J_z(p|_{\mathcal{T}_z})$ and $\bar p_z:=\fint_{\mathcal{T}_z}p\,dx$. Applying (4.3), we deduce

$$\begin{aligned}\|p-\Pi_z(p_z)\|_{L^2(\mathcal{T}_z,\mathbb{M})} &\le \|p-\Pi_z(\bar p_z)\|_{L^2(\mathcal{T}_z,\mathbb{M})}+\|\Pi_z(\bar p_z)-\Pi_z(p_z)\|_{L^2(\mathcal{T}_z,\mathbb{M})}\\ &\le (C_{\rm pf}+C_{\rm j})\,|h\,p|_{H^1(\mathcal{T}_z,\mathbb{M})}.\end{aligned}$$

Therefore

$$\begin{aligned}\|p-Jp\|^2_{L^2(\Omega,\mathbb{M})} &= \int_\Omega\Big\|\sum_{z\in\mathcal{N}}\varphi_z\,(p-\Pi_z p_z)\Big\|^2_{\mathbb{M}}\,dx\le\sum_{z\in\mathcal{N}}\int_{\mathcal{T}_z}\|p-\Pi_z p_z\|^2_{\mathbb{M}}\,dx\\ &\le\sum_{z\in\mathcal{N}}(C_{\rm pf}+C_{\rm j})^2\,|h\,p|^2_{H^1(\mathcal{T}_z,\mathbb{M})}\le 3\,(C_{\rm pf}+C_{\rm j})^2\,|h\,p|^2_{H^1(\mathcal{T},\mathbb{M})}.\end{aligned}$$

Now suppose that $\mathcal{T}$ is shape-regular. Then

$$\begin{aligned}\|h^{-1}(p-Jp)\|^2_{L^2(\Omega,\mathbb{M})} &= \int_\Omega\Big\|\sum_{z\in\mathcal{N}}\varphi_z\,h^{-1}(p-\Pi_z p_z)\Big\|^2_{\mathbb{M}}\,dx\le\sum_{z\in\mathcal{N}}\sum_{T\in\mathcal{T}_z}\|h^{-1}(p-\Pi_z p_z)\|^2_{L^2(T,\mathbb{M})}\\ &\le\sum_{z\in\mathcal{N}}C_{\rm t}^2\,(C_{\rm pf}+C_{\rm j})^2\,|p|^2_{H^1(\mathcal{T}_z,\mathbb{M})}\le 3\,C_{\rm t}^2\,(C_{\rm pf}+C_{\rm j})^2\,|p|^2_{H^1(\mathcal{T},\mathbb{M})}\end{aligned}\tag{4.6}$$

completes the proof. □

Local approximation operators are stable in the $H^1(\Omega)$-norm.

THEOREM 4.3. *Let J be a local approximation operator satisfying* (4.2) *and* (4.4) *and let $\mathcal{T}$ be shape-regular with $C_{\rm sr}$ like in Theorem 3.2. Then for all $p\in V$*

$$|Jp|_{H^1(\Omega,\mathbb{M})}\le\sqrt{3/2}\,C_{\rm sr}\,C_{\rm t}\,(C_{\rm pf}+C_{\rm j})\,|p|_{H^1(\mathcal{T},\mathbb{M})}.$$

PROOF. Since $\sum_{z\in\mathcal{N}}\varphi_z = 1$ on Ω it follows that $\sum_{z\in\mathcal{N}}\nabla\varphi_z$ vanishes. Thus

$$|Jp|^2_{H^1(\Omega,\mathbb{M})} = \int_\Omega \sum_k \Big\| \sum_{z\in\mathcal{N}} (\Pi_z(p_z) - p)\,\partial_k\varphi_z \Big\|^2_{\mathbb{M}} dx \le 3\int_\Omega \sum_{k,z} \|\Pi_z(p_z) - p\|^2_{\mathbb{M}}\,(\partial_k\varphi_z)^2\,dx$$
$$\le 3\max_{x,z}\Big(\sum_k (h\,\partial_k\varphi_z)^2\Big) \sum_{z\in\mathcal{N}} \|h^{-1}\,(\Pi_z(p_z) - p)\|^2_{L^2(\mathcal{T}_z,\mathbb{M})}.$$

To compute $\sum_k (h\,\partial_k\varphi_z)^2$ we can assume without loss of generality that T has the vertices $(0,0)$, $(0,\xi_1)$ and (η_2,ξ_2) and that $\varphi_z(0,0) = 1$ and $\varphi_z(0,\xi_1) = \varphi_z(\eta_2,\xi_2) = 0$. Then

$$\varphi_z(x,y) = \frac{\xi_2/\xi_1 - 1}{\eta_2}\,x - \frac{1}{\xi_1}\,y + 1.$$

Since $|T| = |\xi_1\,\eta_2|/2$, it follows that

$$\sum_k (h\,\partial_k\varphi_z)^2 = \Big(\Big(\frac{h\,\xi_2/\xi_1 - h}{\eta_2}\Big)^2 + \Big(\frac{h}{\xi_1}\Big)^2\Big)\frac{|\xi_1\eta_2|^2}{4\,|T|^2} = \frac{h^2\,((\xi_2-\xi_1)^2+\eta_2^2)}{4\,|T|^2} \le \frac{h_T^4}{2\,|T|^2} \le \frac{C_{\mathrm{sr}}^2}{2}.$$

We already demonstrated in (4.6) that

$$\sum_{z\in\mathcal{N}} \|h^{-1}\,(\Pi_z(p_z) - p)\|^2_{L^2(\mathcal{T}_z,\mathbb{M})} \le 3\,C_{\mathrm{t}}^2\,(C_{\mathrm{pf}} + C_{\mathrm{j}})^2\,|p|^2_{H^1(\mathcal{T},\mathbb{M})},$$

which concludes the proof. □

For shape-regular triangulations $\mathcal{T}$ there is a constant C_{tr} such that for all $T\in\mathcal{T}$ and $p\in H^1(T,\mathbb{M})$

$$\|v\|^2_{L^2(\partial T,\mathbb{M})} \le C_{\mathrm{tr}}(\|h^{-2}\,v\|^2_{L^2(T,\mathbb{M})} + |v|^2_{H^1(T,\mathbb{M})}),$$

cf. [**11**]. Together with the stability of J in $H^1(\Omega)$ this implies that the approximation of p by Jp is on the elemental boundaries of order one half.

THEOREM 4.4. *Let J be a local approximation operator satisfying* (4.2) *and* (4.4) *and let $\mathcal{T}$ be shape-regular with C_{sr} like in Theorem 3.2. Then for all $p\in V$*

$$\sum_{T\in\mathcal{T}} \|h_E^{-1}\,(p - Jp)\|^2_{L^2(\partial T,\mathbb{M})} \le C_{\mathrm{tr}}(1 + C_{\mathrm{p}}^2)(2 + 3\,C_{\mathrm{sr}}^2\,C_{\mathrm{t}}^2\,(C_{\mathrm{pf}} + C_{\mathrm{j}})^2)\,|p|^2_{H^1(\mathcal{T},\mathbb{M})}. \tag{4.7}$$

PROOF. The chain of inequalities

$$\sum_{T\in\mathcal{T}} \|h_T^{-1}\,(p - Jp)\|^2_{L^2(\partial T,\mathbb{M})} \le C_{\mathrm{tr}}^2 \sum_{T\in\mathcal{T}} \|h^{-2}(p - Jp)\|^2_{L^2(T,\mathbb{M})} + |p - Jp|^2_{H^1(T,\mathbb{M})}$$
$$\le C_{\mathrm{tr}}^2(1 + C_{\mathrm{p}}^2) \sum_{T\in\mathcal{T}} |p - Jp|^2_{H^1(T,\mathbb{M})}$$
$$\le C_{\mathrm{tr}}^2(1 + C_{\mathrm{p}}^2)(2|p|_{H^1(\mathcal{T},\mathbb{M})} + 2|Jp|_{H^1(\Omega,\mathbb{M})})^2$$
$$\le C_{\mathrm{tr}}^2(1 + C_{\mathrm{p}}^2)(2 + 3\,C_{\mathrm{sr}}^2\,C_{\mathrm{t}}^2\,(C_{\mathrm{pf}} + C_{\mathrm{j}})^2)\,|p|^2_{H^1(\mathcal{T},\mathbb{M})}$$

shows (4.7). □

5. Reliability

While up to now we have not assumed that p and p_h are related to each other in a particular way, in applications these two functions are connected by an underlying boundary value problem and by the chosen numerical method.

In this section we focus on the proof of reliability of conforming finite element methods for elliptic problems. The analogous results for nonconforming methods and for saddle-point problems are deduced, for instance, in [**15**].

Applying the Einstein summation convention, we consider differential equations of the form

$$-\partial_k\,\mathbb{C}_{ijkl}\,\partial_i u_j = f_l, \tag{5.1}$$

with $u = (u_1, u_2) \in H^1(\Omega, \mathbb{R}^2)$, $f \in L^2(\Omega, \mathbb{R}^2)$ and $\mathbb{C} \in W^{1,\infty}(\Omega, (\mathbb{R}^{2\times 2})^{2\times 2})$. We require that $\mathbb{C}$ is symmetric and uniformly positive definite on $\mathbb{R}^{2\times 2}$; that is $\mathbb{C}_{ijkl} = \mathbb{C}_{klij}$ and there are two positive constants μ and $\dot\mu$ such that

$$\mu\, v_{ij}\, v_{ij} \le v_{ij}\, \mathbb{C}_{ijkl}(x)\, v_{kl} \le \dot\mu\, v_{ij}\, v_{ij},$$

for all $x \in \Omega$ and $v \in \mathbb{R}^{2\times 2}$. Denoting by $\lambda(x)$ the smallest and by $\dot\lambda(x)$ the largest eigenvalue of $\mathbb{C}$, we may select $\mu = \min_{x\in\overline\Omega} \lambda(x)$ and $\dot\mu := \max_{x\in\overline\Omega} \dot\lambda(x)$.

In the context of structural mechanics $\mathbb{C}$ is interpreted as elasticity tensor. For these applications the error control on the flux $p - p_h$, where $p_{ij} := \mathbb{C}_{ijkl}\partial_i u_j$ and $(p_h)_{ij} := \mathbb{C}_{ijkl}\partial_i (u_h)_j$, is often more relevant than the error control on u. Since $\mathbb{C}$ is, in general, not piecewise linear, p_h might not be contained in $\mathcal{P}_h$ as defined in Section 2, a property we do not require for the proof of reliability of $\eta_{\mathcal{E}}$. Towards the end of the section we comment on the condition $p_h \in \mathcal{P}_h$ again.

We define the residual of the finite element method as $r_l := -\partial_k\, \mathbb{C}_{ijkl}(p - p_h)_{ij} = f_l - \partial_k\, \mathbb{C}_{ijkl}(p_h)_{ij}$.

THEOREM 5.1. *Suppose we have the Galerkin orthogonality*

$$\int_\Omega (p_{kl} - (p_h)_{kl})\, \partial_k (J(u - u_h))_l\, dx = 0, \tag{5.2}$$

where J is a suitable local approximation operator. Then

$$\|p - p_h\|_{L^2(\Omega,\mathbb{M})} \le \sqrt{3}\, C_{\rm t}\, (C_{\rm pf} + C_{\rm j})\, \frac{\dot\mu}{\mu}\, \|h\, r\|_{L^2(\Omega,\mathbb{M})} + C_{\mathcal{E}}\, \frac{\dot\mu}{\mu}\, \eta_{\mathcal{E}}, \tag{5.3}$$

where

$$C_{\mathcal{E}} := C_{\rm tr}(1 + C_{\rm p}^2)(2 + 3\, C_{\rm sr}^2\, C_{\rm t}^2\, (C_{\rm pf} + C_{\rm j})^2)\, \max_{T\in\mathcal{T}} \max_{\substack{E\in\mathcal{E}\\ E\subset T}} \frac{h_T}{h_E}.$$

PROOF. Let $\zeta := u - u_h$. Using that $p \in H^1(\Omega, \mathbb{R}^{2\times 2})$, we calculate

$$\begin{aligned}
&\int_\Omega (p_{kl} - (p_h)_{kl})\, \partial_k \zeta_l\, dx = \int_\Omega (p_{kl} - (p_h)_{kl})\, \partial_k (\zeta_l - (J\zeta)_l)\, dx \\
&= \int_\Omega -\partial_k (p_{kl} - (p_h)_{kl})\, (\zeta_l - (J\zeta)_l)\, dx + \sum_{T\in\mathcal{T}} \int_{\partial T} (p_{kl} - (p_h)_{kl})\, (\zeta_l - (J\zeta)_l)\, \nu_k\, dx \\
&= \int_\Omega r_l\, (\zeta_l - (J\zeta)_l)\, dx + \sum_{E\in\mathcal{E}} \int_E [p_h]_{kl}\, (\zeta_l - (J\zeta)_l)\, (\nu_E)_k\, dx \\
&= \|h\, r\|_{L^2(\Omega,\mathbb{M})}\, \|h^{-1}\, (\zeta - J\zeta)\|_{L^2(\Omega)} + \eta_{\mathcal{E}} \Big(\sum_{E\in\mathcal{E}} h_E^{-1}\, \|\zeta - J\zeta\|_{L^2(E)}^2 \Big)^{1/2} \\
&\le \|h\, r\|_{L^2(\Omega,\mathbb{M})}\, \sqrt{3}\, C_{\rm t}\, (C_{\rm pf} + C_{\rm j})\, |\zeta|_{H^1(\mathcal{T},\mathbb{M})} + \eta_{\mathcal{E}}\, C_{\mathcal{E}}\, |\zeta|_{H^1(\mathcal{T},\mathbb{M})}.
\end{aligned}$$

The constant $C_{\mathcal{E}}$ contains besides the coefficients in (4.7) a factor for the transition from h_T to the side length h_E. Assuming that $|\zeta|_{H^1(\Omega,\mathbb{M})} = |\zeta|_{H^1(\mathcal{T},\mathbb{M})} \neq 0$,

$$\|p - p_h\|_{L^2(\Omega,\mathbb{M})} \le \frac{1}{\mu} \Big(\int_\Omega (p_{kl} - (p_h)_{kl})\, \partial_k \zeta_l\, dx \Big)^{1/2} \le \frac{\dot\mu}{\mu}\, \frac{\int_\Omega (p_{kl} - (p_h)_{kl})\, \partial_k \zeta_l\, dx}{|\zeta|_{H^1(\Omega,\mathbb{M})}}.$$

Combining both bounds proves the theorem. □

Condition (5.2) can, for instance, be satisfied for the standard P_1 finite element method with essential Dirichlet boundary conditions. If u_h coincides with u at $\mathcal{N}_\Gamma$ then, also in the case of non-homogeneous boundary conditions, the term $J(u - u_h)$ vanishes if J is defined as nodal interpolation operator. If natural or mixed boundary conditions are imposed, additional terms can arise, for details we refer to [**15**].

We use (5.3) to show the reliability of $\eta_{\mathcal{E}}$. Recall that to apply $\eta_{\mathcal{E}}$ for error control, we need to express $C_{\rm rel}$ and h.o.t.$_{\rm rel}$ defined in the introduction in terms of computable quantities. We do

so with help of *patchwise data oscillations*

$$\operatorname{osc}(r;\mathcal{T}) := \Big(\sum_{z\in\mathcal{N}_\Omega} \operatorname{diam}(\mathcal{T}_z)^2 \Big\| r - \fint_{\mathcal{T}_z} r dx \Big\|^2_{L^2(\mathcal{T}_z,\mathbb{M})} \Big)^{1/2}.$$

Observe that the summation ranges only over the internal nodes. Our next step is to bound r locally in terms of $\eta_{\mathcal{E}}$ and $\operatorname{osc}(r;\mathcal{N})$. The choice of the numerical method is again incorporated into the analysis by (5.5) in form of a Galerkin orthogonality, which implicitly leads to the assumption that

$$\Big\{ \sum_{z\in\mathcal{N}_\Omega} a_z\varphi_z : a_z \in \mathbb{R} \Big\} \tag{5.4}$$

is a subset of the approximation space of the finite element method. Observe that we require in the next theorem $p \in H^1(\Omega,\mathbb{R}^{2\times 2})$.

THEOREM 5.2. *Let $z \in \mathcal{N}_\Omega$ be an interior node. Suppose that $p \in H^1(\Omega,\mathbb{R}^{2\times 2})$ and that*

$$\int_\Omega f\cdot\varphi_z\,dx = \int_\Omega p_h\cdot \operatorname{div}\varphi_z\,dx. \tag{5.5}$$

Then for the residual r we have the local bound

$$\|h\,r\|^2_{L^2(\mathcal{T}_z,\mathbb{M})} \le (1+\sqrt{8}/27\,|\mathcal{T}_z|)\,\operatorname{osc}(r;\mathcal{T}_z)^2 + \sqrt{8}/27\,\|h\|^2_{L^\infty(\mathcal{T}_z)} \Big(\sum_{E\in\mathcal{E}_z} \sqrt{h_E}\,\|[p_h]\cdot\nu_E\| \Big)^2. \tag{5.6}$$

PROOF. We denote the $L^2(\Omega,\mathbb{M})$ scalar product by $\langle\cdot,\cdot\rangle_{\Omega,\mathbb{M}}$. Using that $[p]|_E = 0$ for $E\in\mathcal{E}_z$ and (5.5) we conclude that

$$\begin{aligned}\langle r_l,\varphi_z\rangle_{\Omega,\mathbb{R}} &= \langle (p-p_h)_{kl}, \partial_k\varphi_z\rangle_{\Omega,\mathbb{R}} + \sum_{E\in\mathcal{E}_z}\int_E [p-p_h]_{kl}\,\varphi_z\,(\nu_E)_k\,ds \\ &= \sum_{E\in\mathcal{E}_z}\int_E [p_h]_{kl}\,\varphi_z\,(\nu_E)_k\,ds \le \sum_{E\in\mathcal{E}_z}\frac{\sqrt{h_e}}{3}\,\|[p_h]\cdot\nu_E\|_{L^2(E)}.\end{aligned}$$

The last bound followed from the Cauchy-Schwarz inequality. We denote the identity matrix by I. For the next computation we take into account that $\bar r_z$ and $(\varphi_z-\bar\varphi_z)I$ are $L^2(\Omega,\mathbb{M})$-perpendicular to each other, where $\bar r_z := \fint_{\mathcal{T}_z} r\,dx$ and $\bar\varphi_z := \fint_{\mathcal{T}_z}\varphi_z\,dx$. With $1 = \bar\varphi_z\langle I,I\rangle_{\Omega,\mathbb{M}}/\langle\varphi_z\,I,I\rangle_{\Omega,\mathbb{M}}$,

$$\begin{aligned}\|\bar r_z\|^2_{L^2(\mathcal{T}_z,\mathbb{M})} &= \frac{\langle r,I\rangle^2_{\Omega,\mathbb{M}}}{\langle I,I\rangle_{\Omega,\mathbb{M}}} = \frac{\langle I,I\rangle_{\Omega,\mathbb{M}}}{\langle\varphi_z\,I,I\rangle^2_{\Omega,\mathbb{M}}}\,\langle r,\varphi_z\,I-(\varphi_z\,I-\bar\varphi_z\,I)\rangle^2_{\Omega,\mathbb{M}} \\ &= \frac{\langle I,I\rangle_{\Omega,\mathbb{M}}}{\langle\varphi_z\,I,I\rangle_{\Omega,\mathbb{M}}}(\langle r,\varphi_z\,I\rangle_{\Omega,\mathbb{M}} - \langle r-\bar r_z,(\varphi_z-\bar\varphi_z)\,I\rangle_{\Omega,\mathbb{M}})^2.\end{aligned}$$

From Cavalieri's principle it follows that $\langle\varphi_z\,I,I\rangle_{\Omega,\mathbb{M}} = \sqrt{2}\,\|\varphi_z\|_{L^1(\mathcal{T}_z)} = \sqrt{2}\,|\mathcal{T}_z|/3$. We calculate that $\|(\varphi_z-\bar\varphi_z)\,I\|^2_{L^2(\Omega,\mathbb{M})} = 2\,|\mathcal{T}_z|/18$. Consequently,

$$\begin{aligned}\|\bar r_z\|^2_{L^2(\mathcal{T}_z,\mathbb{M})} &\le \frac{\langle I,I\rangle_{\Omega,\mathbb{M}}}{\langle\varphi_z\,I,I\rangle_{\Omega,\mathbb{M}}}\,(2\,\langle r,\varphi_z\,I\rangle^2_{\Omega,\mathbb{M}} + 2\,\|r-\bar r_z\|^2_{L^2(\mathcal{T}_z,\mathbb{M})}\,\|(\varphi_z-\bar\varphi_z)\,I\|^2_{L^2(\mathcal{T}_z,\mathbb{M})}) \\ &= \sqrt{8}/3\,\langle r,\varphi_z\,I\rangle^2_{\Omega,\mathbb{M}} + \sqrt{8}/27\,|\mathcal{T}_z|\,\|r-\bar r_z\|^2_{L^2(\mathcal{T}_z,\mathbb{M})}.\end{aligned}$$

Therefore,

$$\begin{aligned}\|h\,r\|^2_{L^2(\mathcal{T}_z,\mathbb{M})} &\le \|h\|^2_{L^\infty(\mathcal{T}_z)}\,\|r\|^2_{L^2(\mathcal{T}_z,\mathbb{M})} = \|h\|^2_{L^\infty(\mathcal{T}_z)}\,(\|r-\bar r_z\|^2_{L^2(\mathcal{T}_z,\mathbb{M})} + \|\bar r_z\|^2_{L^2(\mathcal{T}_z,\mathbb{M})}) \\ &\le \|h\|^2_{L^\infty(\mathcal{T}_z)}\,(1+\sqrt{8}/27\,|\mathcal{T}_z|)\,\|r-\bar r_z\|^2_{L^2(\mathcal{T}_z,\mathbb{M})} + \sqrt{8}/3\,\|h\|^2_{L^\infty(\mathcal{T}_z)}\,\langle r,\varphi_z\,I\rangle^2_{\Omega,\mathbb{M}} \\ &\le (1+\sqrt{8}/27\,|\mathcal{T}_z|)\,\operatorname{osc}(r;\mathcal{T}_z)^2 + \sqrt{8}/3\,\|h\|^2_{L^\infty(\mathcal{T}_z)}\Big(\sum_{E\in\mathcal{E}_z}\frac{\sqrt{h_e}}{3}\,\|[p_h]\cdot\nu_E\|_{L^2(E)}\Big)^2.\end{aligned}$$

Notice that only the data oscillations over the subgrid $\mathcal{T}_z$ are needed. □

We remark that by a Poincaré inequality we have for $r \in H^1(\Omega,\mathbb{M})$, i.e. $f \in H^1(\Omega,\mathbb{M})$,

$$\operatorname{osc}(r;\mathcal{T}_z) \leq C_{\mathrm{p}}\ \operatorname{diam}(\mathcal{T})^2\, \|\nabla r\|_{L^2(\mathcal{T}_z,\mathbb{M})}, \tag{5.7}$$

possibly by enlarging C_{p}. Thus the data oscillations are higher-order terms in the sense of (1.1), provided the residuum of the numerical method is bounded from above, independently of h.

We assume that the patches $\mathcal{T}_z$ of internal nodes z cover Ω. This allows us to derive an *a posteriori* bound without explicitly referring to the boundary conditions. In principle when computing the reliability constant C_{rel} only in terms of the internal nodes one should, for the sharpness of the estimate, take into account that while interior elements are always covered by three patches $\mathcal{T}_z$, $z \in \mathcal{N}_\Omega$, at the boundary there are less patches covering the elements. However, for the sake of simplicity we do not make this distinction in the now following analysis.

Finally, for the computation of the reliability constant we use the similar structure between $\eta_{\mathcal{E}}$ and the jump terms in (5.6).

THEOREM 5.3. *Let $\mathcal{T}$ be shape-regular. Provided that* (5.2) *and* (5.5) *hold and that patches of interior elements cover Ω, we have the bound*

$$\|p-p_h\|_{L^2(\Omega,\mathbb{M})} \leq C_{\mathrm{rel}}\ \eta_{\mathcal{E}} + \text{h.o.t.}_{\mathrm{rel}}$$

where

$$C_{\mathrm{rel}}^2 := 2\,\frac{\dot\mu^2}{\mu^2}\left(C_{\mathcal{E}}^2 + C_{\mathrm{en}}\,\sqrt{8}/27\,\|h\|_{L^\infty(\Omega)}^2\right),$$

$$\text{h.o.t.}_{\mathrm{rel}}^2 := 18\,(1+\sqrt{8}/27\,\max_{z\in\mathcal{N}_\Omega}|\mathcal{T}_z|)\,\frac{\dot\mu^2}{\mu^2}\,C_{\mathrm{p}}^2\,C_{\mathrm{t}}^2\,(C_{\mathrm{pf}}+C_{\mathrm{j}})^2\ \operatorname{diam}(\mathcal{T})^2\,\|\nabla r\|_{L^2(\Omega,\mathbb{M})}^2.$$

PROOF. Covering Ω with interior patches gives

$$\begin{aligned}
&\|h\,r\|_{L^2(\Omega,\mathbb{M})}^2 \leq \sum_{z\in\mathcal{N}_\Omega}\|h\,r\|_{L^2(\mathcal{T}_z,\mathbb{M})}^2\\
&\leq \sum_{z\in\mathcal{N}_\Omega}(1+\sqrt{8}/27\,|\mathcal{T}_z|)\,C_{\mathrm{p}}^2\ \operatorname{diam}(\mathcal{T})^2\,\|\nabla r\|_{L^2(\Omega,\mathbb{M})}^2 + C_{\mathrm{en}}\,\sqrt{8}/27\,\|h\|_{L^\infty(\mathcal{T}_z)}^2\sum_{E\in\mathcal{E}_z}h_E\,\|[p_h]\cdot\nu_E\|^2\\
&\leq 3\,(1+\sqrt{8}/27\,\max_{z\in\mathcal{N}_\Omega}|\mathcal{T}_z|)\,C_{\mathrm{p}}^2\ \operatorname{diam}(\mathcal{T})^2\,\|\nabla r\|_{L^2(\Omega,\mathbb{M})}^2 + C_{\mathrm{en}}\,\sqrt{8}/27\,\|h\|_{L^\infty(\Omega)}^2\eta_{\mathcal{E}}.
\end{aligned}$$

Combing this bound with (5.3) gives

$$\begin{aligned}
\|p-p_h\|_{L^2(\Omega,\mathbb{M})}^2 &\leq 6\,C_{\mathrm{t}}^2\,(C_{\mathrm{pf}}+C_{\mathrm{j}})^2\,\frac{\dot\mu^2}{\mu^2}\,\|h\,r\|_{L^2(\Omega,\mathbb{M})}^2 + 2\,C_{\mathcal{E}}^2\,\frac{\dot\mu^2}{\mu^2}\,\eta_{\mathcal{E}}^2\\
&\leq 18\,(1+\sqrt{8}/27\,\max_{z\in\mathcal{N}_\Omega}|\mathcal{T}_z|)\,\frac{\dot\mu^2}{\mu^2}\,C_{\mathrm{p}}^2\,C_{\mathrm{t}}^2\,(C_{\mathrm{pf}}+C_{\mathrm{j}})^2\ \operatorname{diam}(\mathcal{T})^2\,\|\nabla r\|_{L^2(\Omega,\mathbb{M})}^2\\
&\quad +2\,\frac{\dot\mu^2}{\mu^2}\left(C_{\mathcal{E}}^2 + C_{\mathrm{en}}\,\sqrt{8}/27\,\|h\|_{L^\infty(\Omega)}^2\right)\eta_{\mathcal{E}}^2,
\end{aligned}$$

which demonstrates the reliability of $\eta_{\mathcal{E}}$. □

If $p_h \in \mathcal{P}_h$, e.g. $\mathbb{C}$ is elementwise linear and $\mathcal{P}_h = P_1(\mathcal{T},\mathbb{M})$, then the reliability of η_A, η_Z and η_M follows from the equivalence proofs in the section on efficiency. Since the equivalence proofs were in essence based on compactness arguments and did not depend on the linearity of p_h one can show $\eta_A \approx \eta_Z \approx \eta_M \approx \eta_{\mathcal{E}}$ in a more general context. That one can regard the adaptation to differential operators with variable coefficients and higher-order polynomial approximation spaces, we refer to [**6**].

6. Applications

This section gives an overview of some applications to the Stokes and Lamé equations and to elastoplastic, obstacle and degenerated problems in which averaging techniques are known to work. The arguments for the reliability and efficiency proof are partly those from the previous section, partly involve new ideas to cover nonconforming methods and saddle-point problems.

6.1. Stokes Equations. The stationary viscous flow inside a bounded volume $\Omega \subset \mathbb{R}^2$ of viscosity $\mu > 0$ is described by the velocity field $u : \Omega \to \mathbb{R}^2$ and the pressure variable $p : \Omega \to \mathbb{R}$. We assume Dirichlet boundary conditions. Then the *Stokes problem* reads: Given $f \in L^2(\Omega)^2$ and $g \in H^{1/2}(\partial\Omega)^2$, find $(u, p) \in H^1(\Omega)^2 \times L^2(\Omega)$ with

$$\begin{aligned} \operatorname{div} \sigma + f &= 0 \quad \text{in } \Omega, \\ \operatorname{div} u &= 0 \quad \text{in } \Omega, \\ u &= g \quad \text{on } \Gamma, \end{aligned}$$

where $\int_\Omega p(x)\,dx = 0$ and $\sigma := 2\mu\,\varepsilon(u) - pI$ with $\varepsilon(u) := (\nabla u + \nabla u^T)/2$. For some of the results below smoother f and g need to be considered. In the literature one often finds the non-symmetric form $\sigma = 2\mu\nabla u - pI$, which is equivalent to the symmetric model if Dirichlet conditions are imposed.

The weak formulation of the above Stokes problem is straightforward and leads to a mixed FEM with unknowns u_h and p_h. The proper choice of finite spaces for the discrete velocities u_h and the discrete pressures p_h is less trivial, particularly for piecewise polynomials of low-order. We point, for instance, to Kouhia and Stenberg who considered u_h in $\mathcal{S}_k \times \mathcal{S}_k^{nc}$ and piecewise constant p_h.

It is known that, for sufficiently fine meshes, unique discrete solutions (u_h, p_h) for which with quasi-optimal *a priori* error bounds hold, cf. [**21**]. *A posteriori* error estimates are studied in [**4, 17**]. In particular, the reliability and efficiency of the estimators

$$\eta_A := \|\sigma_h - A\sigma_h\|_{L^2(\Omega)} \le \eta_M := \min \|\sigma_h - \tau_h\|_{L^2(\Omega)} \quad \text{for } \tau_h \in \mathcal{S}^1(\mathcal{T}).$$

has been verified for the discrete stress field

$$\sigma_h := 2\mu\varepsilon_{\mathcal{T}}(u_h) - p_h I.$$

For proofs and details see [**17**]. In that publication also mixed boundary conditions are examined. A surprising consequence from the mathematical analysis is that the averaging concerns the stress field σ_h only and not the variables $\varepsilon_{\mathcal{T}}(u_h)$ and p_h separately. The numerical examples in [**17**] underline the high accuracy of η_A in this application.

6.2. Linear Elasticity. The small deformations $u : \Omega \to \mathbb{R}^2$ of a 2D elastic body are modelled by the *Navier-Lamé equations*: Given $f \in L^2(\Omega)^2$, $u_D \in H^{1/2}(\Gamma_D)^2$, $g \in L^2(\Gamma_N)^2$, find $u \in H^1(\Omega)^2$ with

$$\begin{aligned} \operatorname{div} \sigma + f &= 0 \qquad \text{in } \Omega, \\ \operatorname{div} u &= 0 \qquad \text{in } \Omega, \\ u &= u_D \quad \text{on } \Gamma_D, \\ \sigma\nu &= g \qquad \text{on } \Gamma_N, \end{aligned}$$

where

$$\begin{aligned} \sigma &:= \mathbb{C}\,\varepsilon(u) := \lambda \operatorname{tr}(\varepsilon(u))I + 2\mu\,\varepsilon(u), \\ \varepsilon(u) &:= (\nabla u + \nabla u^T)/2. \end{aligned}$$

Here Γ_D is a closed non-empty subset of Γ on which Dirichlet boundary conditions are imposed and Γ_N is a relatively open non-empty subset with Neumann boundary conditions. We assume that $\Gamma_D \cup \Gamma_N = \Gamma$. A new aspect is compressibility with the material constant λ which tends to infinity as the Poisson ratio tends to $1/2$ for rubber-like materials. It is known that in the incompressible limit for $\lambda \to \infty$ that the elastic problem turns into the Stokes problem as $\lambda \operatorname{tr}(\varepsilon(u)) \to p$ and $\operatorname{tr}\varepsilon(u) = \operatorname{div}(u) \to 0$. In principle, the linear elastic problem can be discretised by u_h in $\mathcal{S}_1 \times \mathcal{S}_1$. This leads to quasi-optimal convergence in the energy norm,

$$\|\,\mathbb{C}^{-1/2}(\sigma - \sigma_h)\| \le C(\lambda)\, h_{\max}\, \|D^2 u\|_{L^2(\Omega)}$$

for a smooth exact solution u. Therein, the multiplicative constant $C(\lambda)$ might deteriorate if $\lambda \to \infty$. Figure 1 displays a numerical example from [**16**] for three different materials with Poisson ratios $nu = 0.3$, 0.49, and 0.499. For a uniform sequence of meshes on an L-shaped

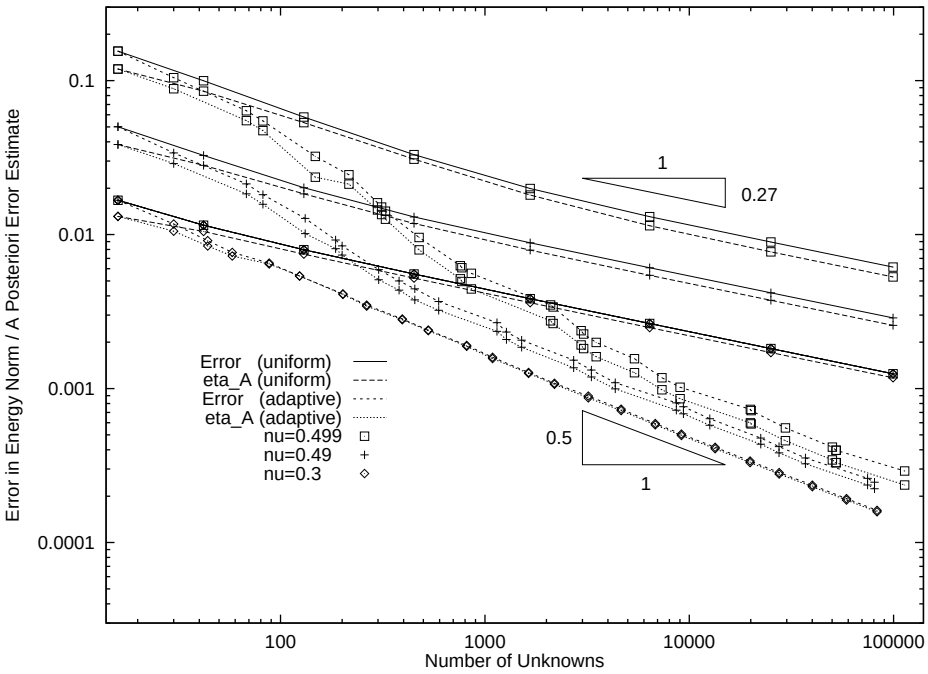

FIGURE 1. Locking in compressible linear elasticity: The energy norm $\| \mathbb{C}^{-1/2}(\sigma - \sigma_h)\|$ and estimators η_A are plotted versus the number of unknowns N for uniform and adapted meshes in the $P_1 \times P_1$ conforming FEM.

domain Ω with known singular exact solution u, we computed discrete solutions $u_h \in \mathcal{S}_1 \times \mathcal{S}_1$ within a standard P_1 FE. The error of the corresponding discrete stress $\sigma_h := \mathbb{C}\,\varepsilon(u_h)$, with the fourth-order material tensor $\mathbb{C}$ and $\sigma = \mathbb{C}\,\varepsilon(u)$, reads, in its energy norm

$$\| \mathbb{C}^{-1/2}(\sigma - \sigma_h)\|^2 := \int_\Omega (\sigma - \sigma_h) : \varepsilon(u - u_h)\, dx.$$

The quantity $\| \mathbb{C}^{-1/2}(\sigma - \sigma_h)\|$ is plotted as a function of the number of degrees of freedom in Figure 1. For uniform meshes we observe a suboptimal convergence rate caused by the singularity of u. The experimental convergence rate appears to be independent of the Poisson ratio nu (i.e. of λ) in contrast to the multiplicative constant $C(\lambda)$. This phenomenon is called locking [**9**]: In Figure 1, the numerical result for a uniform mesh with $N = 10000$ degrees of freedom and $nu = 0.499$ is worse than that for the coarsest mesh with $N = 16$ degrees of freedom for $nu = 0.3$. The situation is even more dramatic for larger and larger $\lambda \to \infty$.

Figure 1 displays three sequences of adaptively refined meshes for $nu = 0.3$, 0.49, and 0.499 as well. The coarse meshes coincide with the results for the uniformly refined ones but improves with a convergence rate larger than 1 until the error is much smaller. Said differently, the effect of the multiplicative constant $C(\lambda)$ is seen in the beginning for $N \leq 100$ and decreases for larger N. Does this indicate the conjecture that adaptivity overcomes locking?

The error control by averaging schemes via

$$\eta_M := \min_{\tau_h \in \mathcal{S}^1_N(\mathcal{T})} \| \mathbb{C}^{-1/2}(\sigma_h - \tau_h)\|_{L^2(\Omega)} \leq \eta_A := \| \mathbb{C}^{-1/2}(\sigma_h - A\sigma_h)\|_{L^2(\Omega)}$$

is displayed in Figure 1 as well. It is proved in [**16**] that $\eta_M \leq \eta_A$ is reliable up to λ-depending constants and, clearly, η_M is efficient with respect to λ-independent constants. As observed in Figure 1, even a very poor finite element solution is estimated very accurately.

The preceding discussion focused on conforming FEM and large errors caused by locking in the incompressible limit $\lambda \to \infty$. More appropriate FEMs can overcome this locking. The first

hint is to use ansatz and test functions which lead to stable FEM for the Stokes problem regarded as the limit problem for incompressibility. The choice of $\mathcal{S}_k \times \mathcal{S}_k^{nc}$ due to Kouhia and Stenberg [**21**] is appropriate for that and leads in [**16**] to robust and accurate estimates. Therein, the finite element schemes as well as their error estimators are highly accurate and λ-independent.

6.3. Elastoplasticity. This section briefly describes the perspectives and limitations of averaging techniques in elastoplastic evolution problems. Therein, a time-discretisation is performed followed by a spatial discretisation in each time-step. Averaging error estimators $\eta_M \leq \eta_A$ for the exact and discrete stress field σ and σ_h, respectively, have the same definition as in Subsection 7.2 for each time step. It can be proved for an implicit time-discretisation that

$$\|\mathbb{C}^{-1/2}(\sigma-\sigma_h)\|^2 \leq \int_\Omega (\sigma-\sigma_h) : \varepsilon(u-u_h)\,dx = \int_\Omega (\sigma-\sigma_h) : \varepsilon(u-u_h-v_h)\,dx \quad \text{for all } v_h \in \mathcal{S}_h \times \mathcal{S}_h.$$

This term is an upper bound for error terms such as the stress error in the energy norm. Moreover, if hardening is present, the stress error controls the displacement error $\|u-u_h\|_{H^1(\Omega)}$ up to hardening-depending multiplicative constants. For details and proofs we refer to [**2, 13, 14**]. Therefore, one can proceed as in linear elasticity to derive reliability and efficiency of $\eta_M \leq \eta_A$. The constant $C_{\rm rel}$, however, depends crucially on the hardening; the estimates are *not* valid (in this form) for perfect plasticity.

It should be emphasized that reliability holds solely for the spatial discretisation; the accumulated error in time is *not* controlled by $\eta_M \leq \eta_A$; the result holds for Hencky materials only. The control of the time-discretisation error appears to be an important open question. However, the numerical results in [**14**] provide numerical evidence that η_A is indeed a very accurate (spatial) error estimator.

6.4. Obstacle Problems. This section briefly highlights the surprising result that for nonlinear variational inequalities, in certain settings, the same averaging estimator

$$\eta_M := \min_{q_h \in \mathcal{Q}_h} \|p_h - q_h\| \leq \eta_A := \|p_h - Ap_h\|$$

is, as for the variational equation, reliable and efficient: An affine obstacle has no substantial influence! We illustrate the situation with a linear model problem. Let K denote the set of admissible deformations,

$$K := \{v \in H^1_0(\Omega) : \, 0 \leq v \text{ almost everywhere in } \Omega\}$$

with $H^1_0(\Omega) := \{v \in H^1(\Omega) : \, v = 0 \text{ on } \partial\Omega\}$. Then the weak form of the obstacle problem reads: Given $f \in L^2(\Omega)$ find $u \in K$ with

$$\int_\Omega \nabla u \cdot \nabla(u-v)\,dx \leq \int_\Omega f(u-v)\,dx \quad \text{for all } v \in K.$$

The FE discretisation replaces K by the discrete version

$$K_h := K \cap P_1(\mathcal{T})$$

and hence determines $u_h \in K_h$ with

$$\int_\Omega \nabla u_h \cdot \nabla(u_h - v_h)\,dx \leq \int_\Omega f(u_h - v_h)\,dx \quad \text{for all } v_h \in K_h.$$

The main difference of the variational inequality and the model example of Section 3 can be expressed by means of residuals $\varrho \in H^{-1}(\Omega)$ and $\varrho_h \in P_1(\mathcal{T})$,

$$\begin{aligned} \varrho(v) &:= \int_\Omega v\, f\,dx - \int_\Omega \nabla u \cdot \nabla v\,dx \quad \text{for all } v \in H^1(\Omega), \\ \varrho_h &:= \sum_{z \in \mathcal{K}} \left(\int_\Omega f\,\varphi_z\,dx - \int_\Omega \nabla u_h \cdot \nabla \varphi_z\,dx \right) \psi_z / \int_\Omega \varphi_z\,dx \quad \in P_1(\mathcal{T}). \end{aligned}$$

It is elementary to verify that the error $e := u - u_h$ in the energy norm reads

$$|e|^2_{1,2} = \int_\Omega f(e-e_h)\,dx - \int_\Omega \nabla u_h \cdot \nabla(e-e_h)\,dx + \int_\Omega \varrho_h\, e\,dx - \varrho(e)$$

for some $e_h := \sum_{z\in\mathcal{N}_\Omega} \left(\int_\Omega e\psi_z\,dx\right) \varphi_z / \int_\Omega \varphi_z\,dx \in P_1(\mathcal{T})$. The approximation error $e - e_h$ can be analyzed as previously in the text. The additional terms with ϱ_h and ϱ reflect the variational inequality. Indeed, one can show that $0 \le \varrho(e)$ and

$$\varrho_h(z) \le 0 = \varrho_h(z)u_h(z) \le u_h(z) \quad \text{for all } z \in \mathcal{N}_\Omega.$$

Hence the arguments of Section 5 lead to the *a posteriori* error estimate

$$|e|_{1,2}^2 \le C\eta_M^2 + \text{h.o.t.}(f) - \int_\Omega \varrho_h\, u_h\, dx.$$

The last term can be analyzed further. Indeed, $\varrho_h u_h$ vanishes on an element $T \in \mathcal{T}$ or, at least, $\varrho_h(a) < 0 = u_h(a) = \varrho_h(b) < u_h(b)$ for two nodes a and b of T. Inverse inequalities based on $\varrho_h \le 0 \le u_h$ yield a bound of $\|\varrho_h u_h\|_{L^1(T)}$ in terms of the mesh-size, $|\varrho_h|_{1,2}$, and $\|\nabla u_h - A(\nabla u_h)\|_2$. The details and proofs can be found in [**5**]. The final result reads

$$|e|_{1,2} \le C_{\text{rel}}\eta_M + \text{h.o.t.}(f).$$

For non-affine obstacles and nonconforming discretisations (i.e. $K_h \not\subset K$) some consistency terms arise and may dominate the upper bound, cf. [**5**]. Numerical results in [**5**] provide empirical evidence for a surprisingly high accuracy of η_A.

6.5. Degenerate Problems. The preceding examples concerned uniformly convex minimization problems on affine or convex subsets. The p-Laplacian is a first nonlinear equation with less strong convexity which requires a particular analysis. This is based on an appropriate quasi-norm, a metric that depends on the exact or discrete solution. The techniques of Section 5 and 6, however, can be adopted to this setting and then yield reliable and efficient error estimators [**18**].

The situation is even more difficult and essentially open for convexified problems where the energy minimization functional is not strictly convex. Very much as a surprise came numerical evidence in a 2-well benchmark example allowing for microstructures that η_A yields an accurate stress error estimation [**12**].

References

1. Mark Ainsworth and J. Tinsley Oden, *A posteriori error estimation in finite element analysis*, Pure and Applied Mathematics (New York), Wiley-Interscience [John Wiley & Sons], New York, 2000.
2. Jochen Alberty and Carsten Carstensen, *Numerical analysis of time-depending primal elastoplasticity with hardening*, SIAM J. Numer. Anal. **37** (2000), no. 4, 1271–1294.
3. Ivo Babuška and Theofanis Strouboulis, *The finite element method and its reliability*, Numerical Mathematics and Scientific Computation, The Clarendon Press Oxford University Press, New York, 2001.
4. Weizhu Bao and John W. Barrett, *A priori and a posteriori error bounds for a nonconforming linear finite element approximation of a non-Newtonian flow*, RAIRO Modél. Math. Anal. Numér. **32** (1998), no. 7, 843–858.
5. S. Bartels and C. Carstensen, *Averaging techniques yield reliable a posteriori finite element error control for obstacle problems*, Numer. Math. **99** (2004), no. 2, 225–249.
6. Sören Bartels and Carsten Carstensen, *Each averaging technique yields reliable a posteriori error control in FEM on unstructured grids. II. Higher order FEM*, Math. Comp. **71** (2002), no. 239, 971–994.
7. Roland Becker and Rolf Rannacher, *An optimal control approach to a posteriori error estimation in finite element methods*, Acta Numer. **10** (2001), 1–102.
8. Peter Binev, Wolfgang Dahmen, and Ron DeVore, *Adaptive finite element methods with convergence rates*, Numer. Math. **97** (2004), no. 2, 219–268.
9. Dietrich Braess, *Finite elements*, second ed., Cambridge University Press, Cambridge, 2001, Theory, fast solvers, and applications in solid mechanics, Translated from the 1992 German edition by Larry L. Schumaker.
10. C. Carstensen, *All first-order averaging techniques for a posteriori finite element error control on unstructured grids are efficient and reliable*, Math. Comp. **73** (2004), no. 247, 1153–1165.
11. C. Carstensen and S. A. Funken, *Constants in Clément-interpolation error and residual based a posteriori error estimates in finite element methods*, East-West J. Numer. Math. **8** (2000), no. 3, 153–175.
12. C. Carstensen and K. Jochimsen, *Adaptive finite element methods for microstructures? Numerical experiments for a 2-well benchmark*, Computing **71** (2003), no. 2, 175–204.
13. Carsten Carstensen, *Numerical analysis of the primal problem of elastoplasticity with hardening*, Numer. Math. **82** (1999), no. 4, 577–597.
14. Carsten Carstensen and Jochen Alberty, *Averaging techniques for reliable a posteriori FE-error control in elastoplasticity with hardening*, Comput. Methods Appl. Mech. Engrg. **192** (2003), no. 11-12, 1435–1450.

15. Carsten Carstensen and Sören Bartels, *Each averaging technique yields reliable a posteriori error control in FEM on unstructured grids. I. Low order conforming, nonconforming, and mixed FEM*, Math. Comp. **71** (2002), no. 239, 945–969.
16. Carsten Carstensen and Stefan A. Funken, *Averaging technique for FE—a posteriori error control in elasticity. II. λ-independent estimates*, Comput. Methods Appl. Mech. Engrg. **190** (2001), no. 35-36, 4663–4675.
17. ———, *A posteriori error control in low-order finite element discretisations of incompressible stationary flow problems*, Math. Comp. **70** (2001), no. 236, 1353–1381.
18. Carsten Carstensen, W. Liu, and N. Yan, *A posteriori error estimators based on gradient recovery for finite element approximation of the p-laplacian*, Report UKC/IMS (2001), no. 01/44.
19. Willy Dörfler, *A convergent adaptive algorithm for Poisson's equation*, SIAM J. Numer. Anal. **33** (1996), no. 3, 1106–1124.
20. Willy Dörfler and Ricardo H. Nochetto, *Small data oscillation implies the saturation assumption*, Numer. Math. **91** (2002), no. 1, 1–12.
21. Reijo Kouhia and Rolf Stenberg, *A linear nonconforming finite element method for nearly incompressible elasticity and Stokes flow*, Comput. Methods Appl. Mech. Engrg. **124** (1995), no. 3, 195–212.
22. Ricardo H. Nochetto, *Removing the saturation assumption in a posteriori error analysis*, Istit. Lombardo Accad. Sci. Lett. Rend. A **127** (1993), no. 1, 67–82 (1994).
23. O. C. Zienkiewicz and J. Z. Zhu, *A simple error estimator and adaptive procedure for practical engineering analysis*, Internat. J. Numer. Methods Engrg. **24** (1987), no. 2, 337–357.

Department of Mathematics, Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Germany

E-mail address: cc@mathematik.hu-berlin.de
E-mail address: jensenm@mathematik.hu-berlin.de